Introduction

According to the data of the Indian Council of Medical Research, with prostate cancer increasingly becoming common among men, particularly in metros, it has moved from being the eighth most common cancer in males in the 1990s to the third rank by 2015 in India. Chronic inflammation is associated with the onset of Prostate Cancer (PCa), where infection plays a major role in both inflammation and cancer development[1]. The prostate gland harbours a relatively higher number of lymphocytes and other immune cells to protect it from invading pathogens[2]. While prostate-specific antigen (PSA) testing followed by trans-urethral resection of the prostate has long been recommended, through this method, PCa detection at an early stage is not always accompanied by an accurate determination of morbidity risk, and therefore mortality might result due to over-treatment in some cases and under-treatment in others. Thus, great attention has been paid to exploring better molecular approaches for the early diagnosis and prognosis of PCa. Recently, a number of next-generation sequencing as well as hybridization-based microarray studies identified a potential microbiome signature associated with PCa development[3,4,5], suspecting it to be the prime factor for inflammation mediated prostate cancer development. Despite these efforts, the pathogenic microbial composition of PCa may vary among populations of different ethnicities across the globe. Given the increasing incidence of PCa patients in India[6], a detailed and comprehensive analysis of the microbial ecosystem coupled with PCa development is of utmost importance.

Apart from prostate cancer, another prevalent cancer in India is oral squamous cell carcinoma (OSCC). Our country has the second-highest number of oral cancer cases. Tobacco consumption is a prime cause of oral cancer, particularly in India. Apart from tobacco, chewing paan containing leaves of piper betel with areca nut, lime, catechu, cinnamon, etc. is a leading source of oral malignancy. As one of the prime territories of microbiome in human body, the oral cavity contains distinct niches with dynamic microbial communities[7]. With advancements in science and technology,researchers from various parts of the world have investigated the association of oral microbial community with various clinical symptoms such as dental caries, periodontal disease, and oral cancer[8,9,10,11,12].So it is very clear that oral microbial ecology is a critical factor in controlling both oral physiology and pathophysiology, but to our surprise, till date no efforts have been made in understanding the oral microbial imbalance during OSCC development among Indian patients, despite having the highest oral cancer incidence.

The advent of next-generation sequencing technologies, for example, 16S rRNA gene amplicon based sequencing, has allowed an affordable and culture-free approach to the identification of overall bacterial composition in cancerous lesions and its effect on the progression of the disease[13]. Herein, we aim to identify specific microbial signatures associated with the pathologic prostate and oral tissue specimens collected from patients in the Eastern region of India. In our studies, we aim to explore PCa and OSCC-associated bacterial composition to find the signature alterations of the bacterial community during PCa and OSCC development and indicate the possibility of utilizing the identified microbiome signature as a prognostic marker of oral malignancies in patients of the Indian subcontinent.

Objectives

- Identification of microbial signature associated with the development of prostate cancer.
- Identification of microbial signature associated with the development of oral cancer.

Methodology

Ethics Statement:

- **Prostate Cancer**

  The study was approved by the Institutional Review Board of Institute of Post Graduate Medical Education & Research (IPGME&R), Kolkata, India. Written informed consent was obtained from all participants.

- **Oral Cancer**

  The study was approved by the Institutional Ethics Committee for Human Research, Indian Statistical Institute, Kolkata, India. Written informed consent was obtained from all participants.

All methods in these studies were performed in accordance with the ethical principles founded in the Declaration of Helsinki.

Sample Information:

- **Prostate Cancer**

  Affected individuals were recruited at Dept. of Urology, IPGME&R, Kolkata. A standard TRUS-guided 18-core prostate biopsy with subsequent histopathological analysis was carried out in those with abnormal serum PSA level (>4.0 ng/ml). All patients were without any prior treatment before surgery. Samples were collected into two cohorts. Cohort-1 (Discovery Cohort) containing 13 BPH and 33 PCa tissue biopsy samples were subjected to 16S rRNA amplicon sequencing. Cohort-2 (Validation Cohort) containing 15 BPH and 16 PCa tissue biopsy samples together with Cohort-1 were used for validation using real-time qPCR analyses.

- **Oral Cancer**

  After the clinical diagnosis of oral squamous cell carcinoma (OSCC) the patients from Dr. R Ahmed Dental College and Hospital, were enlisted for the study. 50 patients were included in the study after confirmation of well-differentiated squamous cell carcinoma from histopathological reports. No participants were on any local or systemic antibiotics prior to sample collection. Tissue samples were collected by incisional and 3 mm punch biopsy sample collection method from both the regions of cancerous lesions (N=50) and the adjoining clinically normal area (matched control, N=50) for each of the 50 patients recruited in this study.

Clinical characteristics and other information related to both types of samples are described in Table 1 and Table 2.

DNA Extraction:

DNA was isolated from the tissue regions using DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer's protocol. The quality and quantity of DNA was determined by the A260/280 ratio using Synergy H1 Multimode Microplate Reader. Approximately 50 ng of genomic DNA from each sample was used for 16S rRNA amplicon sequencing.

16S rRNA Amplicon Sequencing and OTU Assignments:

- **Prostate Cancer**

  For characterization of bacterial taxonomy and species identification, different hypervariable regions of 16S rRNA genes were amplified and sequenced on GeneStudio S5 System using Ion 550 Chip. AmpliSeq Libraries were prepared using an automated Ion Chef system and subsequently placed in the emulsion PCR. The concentration and quality of the amplicons were measured using Agilent 2100 Bioanalyzer.

  To assign taxonomy, each unique sequence was mapped using two 16S rRNA reference databases – Greengenes v13.5 and the Thermo Fisher Scientific in-house MicroSeq 500 v2013.1. Reads were aligned against the databases using MegaBLAST. The expectation value (E-value) was set to 0.01. To assign taxonomy, the minimum alignment percentage was set to a threshold of 90. A read was assigned to a genus only when the identity score of the sequence alignment was at 97% or higher. For species assignment, the minimum percentage was set to 99%.

- **Oral Cancer**

  16S ribosomal RNA (rRNA) amplicon sequencing for metagenomics studies was performed on a MiSeq platform (Illumina) using 2×250 bp chemistry. Clonal libraries for 16S rRNA V3-V4 hypervariable region were prepared using NEBNext Ultra DNA Library preparation. The adapter ligated fragments were indexed by limited PCR cycle to generate final libraries for paired-end sequencing. The concentration of the purified amplicons was measured using Qubit fluorometer and the quality was checked using Agilent 2100 Bioanalyzer. The multiplex amplified libraries were pooled in equimolar concentrations with unique indices, mixed with 15% PhiX control and sequenced using MiSeq reagent kit v2 (Illumina) according to manufacturer's instruction.

  The raw FASTQ sequencing files were further processed FASTQC toolkit. The targeted amplicons were filtered out from other superfluous sequences by detecting the specific conserved region. Forward and reverse reads were stitched together with a minimum overlap of 30 bp and maximum overlap of 250 bp. The taxonomy classification and relative abundance assignments were performed using 'Quantitative Insights into Microbial Ecology' (QIIME v. 1.9.0)[14] pipeline. To confirm the annotation, the resulting OTU sequences were then searched against the Ribosomal Database Project naïve Bayesian classifier[15,16] database. The taxonomy classifications at the phyla, order, family, genera and species level were performed using the GreenGenes and RDP databases.

Diversity and Bacterial Enrichment Analyses:

MicrobiomeAnalyst[17]was used for statistical analysis. The α-diversity indexes including observed OTU numbers, Chao1 index, Simpson index, and Shannon index and the β-diversity – Bray-Curtis dissimilarity measurements were calculated. Evolutionary relation of the genera unique to the malignant samples and the normal were analyzed and a cladogram was generated using Galaxy[18]. The variation as well as the unique bacterial composition in the normal and malignant samples was identified using Random Forest[19] classification analysis within MicrobiomeAnalyst. To estimate β-diversity, un-weighted and weighted UniFrac distances by Bray-Curtis method were calculated from the OTU abundance and utilized in Principal Component Analysis (PCoA) to analyze the unique clustering genera for the normal and malignant tissue samples. The differential abundances of OTUs and specific OTU enrichment between cancer samples and controls were determined using LEfSe based on Kruskal–Wallis H test.

Functional Prediction of Distinct Bacterial Communities:

Functional compositions of the bacterial communities among two different groups were predicted using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt)[21]according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[22].

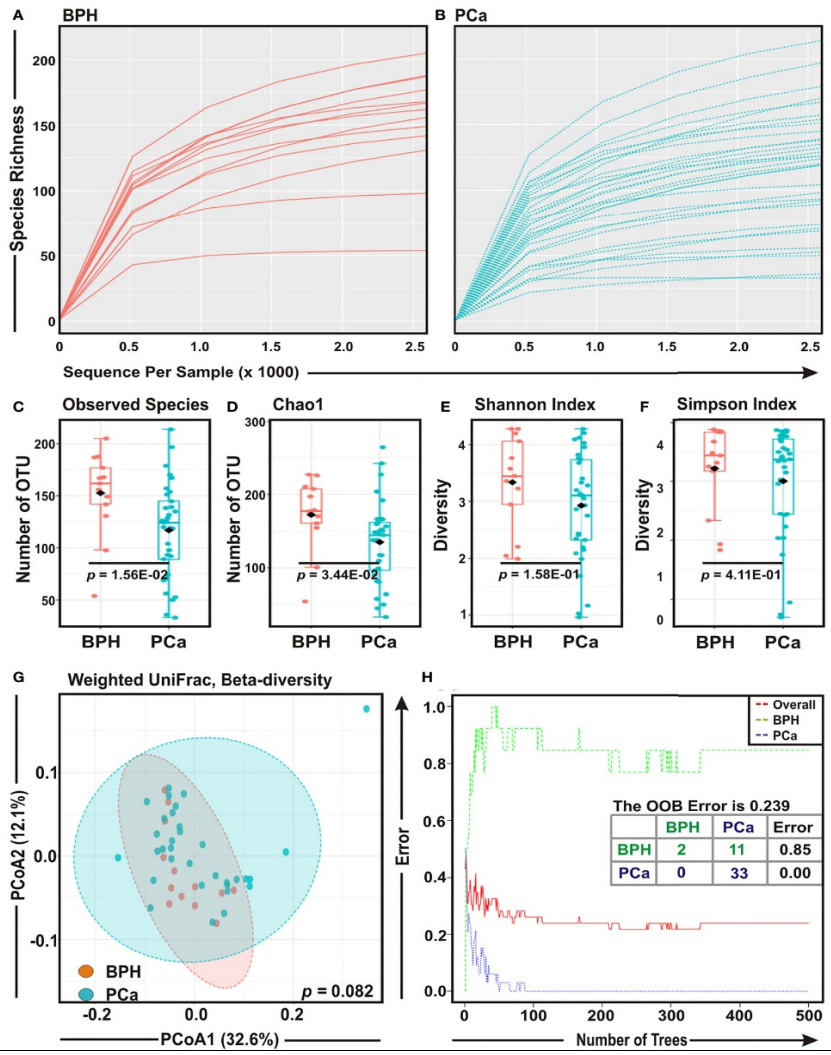

FIGURE 1: (A, B) Rarefaction analysis of bacterial 16S rRNA amplicon. (C–F) Box-Whisker plots of OTU abundance. (G) Principal Component Analysis (PCoA) plot based on weighted UniFrac distance (H) The random forest classification analyses of BPH and PCa samples.

Real-Time PCR Primer Designing and Data Output:

Primer-BLAST tool in National Center for Biotechnology Information (NCBI) database was used to design primers for real-time quantitative PCR (qPCR) analyses. Primers for conserved sequence of bacterial 16S rRNA gene, human Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene and human oncogenic viruses are listed in Table 3.

The optimum primer melting temperature (Tm) was chosen at 60°C and the maximum GC content was kept at 55%. qPCR analysis was performed using iTaq Universal SYBR Green Supermix (BIO-RAD) in CFX Connect Real-Time PCR detection System (BIO-RAD). Each sample was performed in duplicate and calculation was made using a −ΔCT method to quantify relative abundance compared with human genomic GAPDH control. The -ΔCt values of each OSCC samples and their match controls were plotted using GraphPad Prism 8.0.1.

Correlation Analysis Between Microbes:

Correlation among bacteria and oncogenic viruses were analyzed using R software. Spearman's rank test was performed using the ΔCt value of both bacteria and viruses. P-value cutoff was adjusted to 0.05 significance level.

Results

Objective I - Identification of microbial signature associated with the development of prostate cancer.

Sample Characteristics and Sequencing Data Summary of Diseased Prostate Samples:

To understand the changes of microbial compositions associated with prostate cancer (PCa) development, we collected a total of 46 tissue biopsy samples from 13 benign prostatic hyperplasia (BPH) and 33 PCa patients ('Discovery Cohort,' Table 1). The total DNA of each specimen from the 'Discovery Cohort' was extracted and subjected to 16S rRNA sequencing on an GeneStudio S5 System targeting the hypervariable regions of 16S rRNA gene. Additionally, as described later, to validate the sequencing data, we also collected another 31 tissue biopsy samples from 16 BPH and 15 PCa patients ('Validation Cohort,' Table 1).

A total of 35,320,332 raw reads were generated of the 46 samples from the 'Discovery Cohort'. The number of raw sequence reads varied by approximately 10 fold across samples. After quality trimming and chimera checking, 30,641,026 high-quality reads were mapped, while 6,301,956 reads were excluded due to low copy number reads (<10 copy numbers) and 41,312 reads were found to be un-mapped. Subsequently, 24,297,758 mapped reads identified Operational Taxonomic Units (OTUs) at the level of family (>95% similarity index), genus (>97% similarity index), and species (>99% similarity index). Overall, a total of six phylum, 127 genera, and 291 bacterial species were identified from all 46 samples in the Discovery Cohort.

Overall Bacterial Abundance and Diversity Among BPH and PCa Samples:

Rarefaction curves of species richness were plotted for BPH and PCa samples to determine the efficiency of the sequencing process (Figures 1A, B). Most of the samples, though not completely, reached a saturated plateau phase, indicating that the depth of sequencing was sufficient for the diversity analysis (Figures 1A, B). Both Observed Species ($p = 0.015$) and Chao1 index ($p = 0.034$) showed that species richness was significantly decreased in PCa samples (Figures 1C, D). In addition, the diversity estimators for both Shannon index ($p = 0.158$) and Simpson index ($p = 0.411$) indicated a depletion in relative diversity of species composition in PCa samples as compared to BPH samples, although the data were not statistically significant (Figures 1E, F).

To assess the diversity among two groups, we evaluated weighted UniFrac distance matrix subsequently applied in Principal Component Analysis (PCoA) (Figure 1G). PCoA analyses revealed no significant difference ($p = 0.082$) in the bacterial compositions between BPH and PCa samples (Figure 1G), a 'Random Forest' algorithm was applied to further confirm the difference in bacterial community among the BPH and PCa biopsy samples (Figure 1H).

The decision trees identified distinct bacterial composition in two BPH samples and 11 BPH samples exhibited overlapping species with PCa samples (class error: 0.850; Figure 1H). In contrast, all 33 PCa samples demonstrated unique bacterial compositions (Figure 1H).

Taxonomic characterization of Bacterial Compositions Among BPH and PCa Samples:

The bacterial communities associated with the diseased prostate lesions were further analyzed at different taxonomic levels. Six phyla comprised the entire sequences in both groups (Figures 2A, B). In general, Proteobacteria was the most abundant phylum in the prostate microbial ecology (Figures 2A, B). Overall, the results indicated that only Actinobacteria phylum was significantly depleted in PCa samples as compared to the BPH (Figures 2A, B). At the genus level, *Prevotella*, *Cupriavidus*, *Propionibacterium*, *Acinetobacter*, and *Corynebacterium* represented the top

five genera in both samples (Figures 2C, D). To identify the differentially enriched genera within BPH and PCa samples, we employed Linear Discrimination Analysis (LDA) Effect Size (LEfSe) method (Figure 2E).

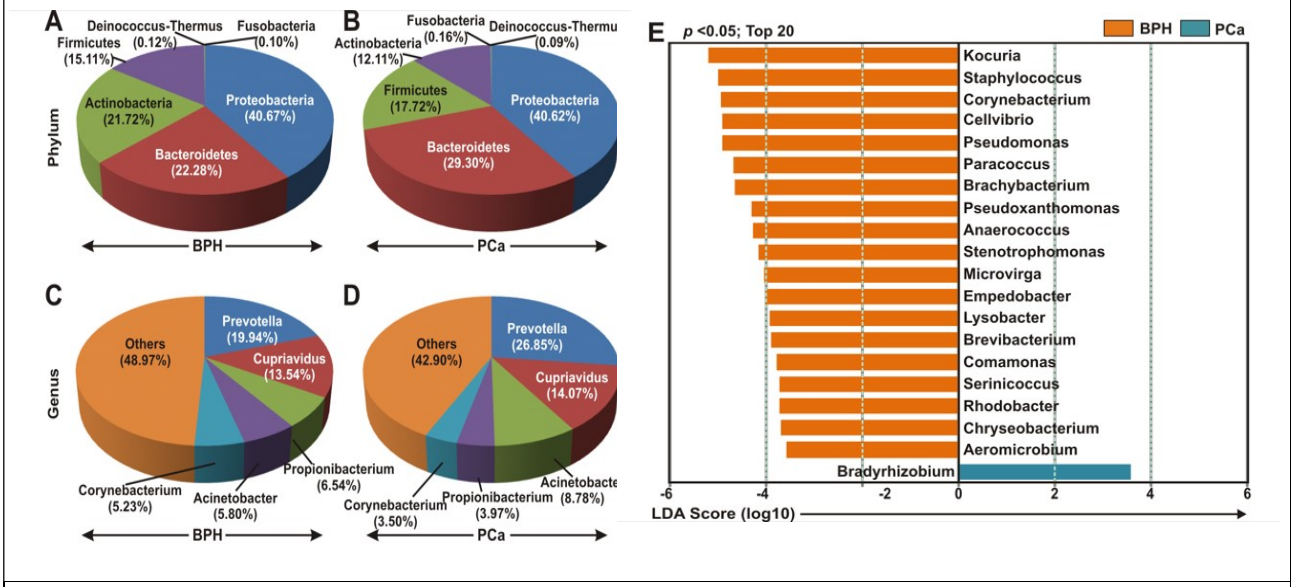

FIGURE 2: (A–D) Relative abundance of bacterial communities at the (A, B) phylum and (C, D) genus levels in (A–C) BPH and (B, D) PCa tissue samples. (E) Association of specific microbiota taxa (top 20 bacterial genera) by LEfSe analyses in BPH and PCa samples.

Validation of Species Identified in 16S rRNA Sequencing by Quantitative Real-Time PCR Analyses in BPH and PCa Samples:

To confirm the 16S rRNA sequencing data, quantitative real-time PCR (qPCR) analyses were performed of top two bacteria in each category – BPH (*K. palustris* and *C. mixtus*) and PCa (*C. taiwanensis* and *Methylobacterium organophilum*), along with three most abundant bacteria (*P. copri*, *C. campinensis*, and *P. acnes*) identified in diseased prostate lesions, using two distinct species-specific primers (Figure 3).

Moreover, in order to further corroborate the results, we used two sample cohorts – 'Cohort-1' (Discovery Cohort) and 'Cohort-2' (Validation Cohort) as described in Table 1. As similar to LEfSe analyses between BPH and PCa biopsy samples, qPCR data further confirmed *Kocuria palustris* and *Cellvibrio mixtus* as BPH specific and *Cupriavidus taiwanensis*, and *Methylobacterium organophilum* as PCa specific bacterial species in both sample cohorts (Figures 3A, D).

Functional Prediction of Altered Microbiome Associated With the PCa Development:

In order to predict and visualize the functional effects resulting from the altered microbial community associated with disease progression in prostate, we employed Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) software. The LEfSe outputs of KEGG pathways shows that pathways such as nitrotoluene degradation, steroid hormone biosynthesis, non-homologous end-joining, and primary bile acid biosynthesis were directly linked to PCa development (Figure 4A). PCoA also demonstrated that the predicted functions of bacterial compositions among BPH and PCa were significantly clustered ($p < 0.05$) (Figure 4B).

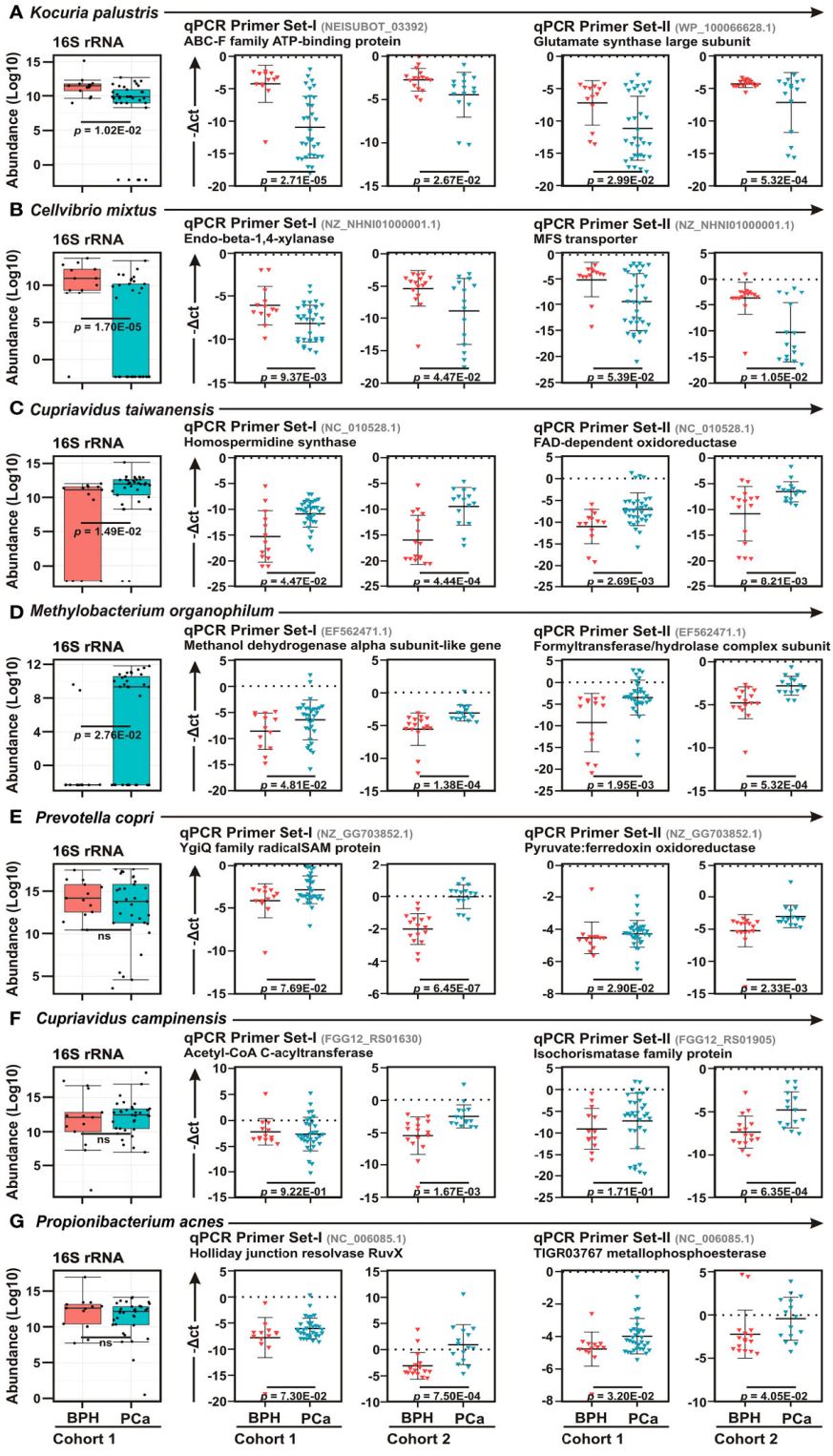

FIGURE 3 Real-time qPCR confirmation of bacterial species among BPH and PCa lesions. (A–G) Box Whisker plots of species enrichment.

Quantitative Real-Time PCR Analyses Detected Strong Association of Multiple Human Tumor Viruses WithPCa Progression:

Studies suggest that a number of human oncogenic viruses including high-risk HPVs and EBV are associated with PCa development. To evaluate the potential involvement of viral etiology in our samples we designed qPCR primers for seven human tumor viruses including EBV, two high risk HPVs – HPV-16 and HPV-18 –, HBV, HTLV-1, HCV, KSHV, and MCPyV along with two more human polyomaviruses – JCV and BKV (Table 3).

qPCR analyses of Cohort-1 with primer set-I against all virus-specific antigens demonstrated that only EBV, HPV-16, HPV-18, and HBV are significantly associated with PCa samples (Figures 5A–D).

The housekeeping gene, human GAPDH gene, was utilized as control in both BPH and PCa samples. A higher negative -ΔCt (average GAPDH Ct value – average target primer Ct value) indicated elevated presence of the virus in the samples. We further validated the association of these four selected tumor viruses in Cohort-2 using a second set of primers (Figures 5A–D and Table 2).

The results clearly demonstrated that all four tumor viruses were significantly associated with the PCa lesions in comparison to BPH samples (Figures 5A–D).

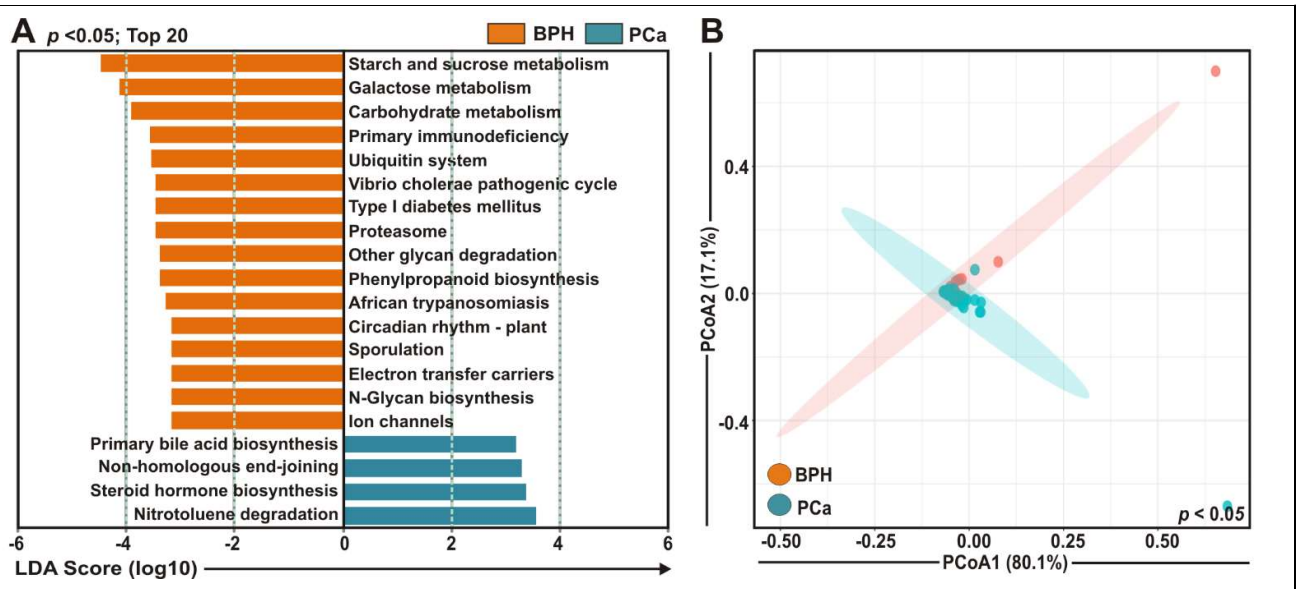

FIGURE 4: (A) LEfSe analyses of potential functional pathways. (B) Principal Component Analysis (PCoA) plot comparing the bacterial functions associated with BPH and PCa tissue biopsy samples.

Co-Occurrence of Tumor Viruses With Microbiome Signature Linked to BPH and PCa Lesions

To further corroborate the connection of these tumor viruses with the identified microbial signature associated with BPH and PCa lesions, the co-occurrence and co-exclusion patterns of EBV, HPV-16, HPV-18, and HBV with the most abundant bacterial species identified in LEfSe and qPCR analyses in each category were further investigated (Figure 5E). Among BPH-specific bacteria, interestingly only K. palustris was positively correlated with EBV (r = 0.446) and as expected negatively correlated with HPV-16 (r = -0.461) and to a lesser extent with HBV (r = -0.263) (Figure 5E). Among PCa-specific bacteria, while *C. taiwanensis* was positively correlated with HPV-16 (r = 0.411) and to lesser extents with HPV-18 (r = 0.328) and HBV (r = 0.246), *M. organophylum* were positively correlated with EBV (r = 0.456) and to lesser extents with HPV-18 (r = 0.325) and HPV-16 (r = 0.261) (Figure 5E).

Objective II - Identification of microbial signature associated with the development of oral cancer.

Subject Characteristics and Oral Microbiota Profiling by 16S rRNA V3-V4 Amplicon Sequencing:

To investigate changes in the oral microbiome associated with OSCC development, we collected 50 pair of OSCC lesions and adjacent healthy tissue samples. The bacterial DNA was isolated from all 50 pair specimens (Table 2), followed by PCR amplification targeting the 16S rRNA V3–V4 hypervariable region. The amplified libraries were sequenced on an Illumina MiSeq platform. A total of 477,411 raw sequences were generated and after quality trimming and chimera checking, 322,656 high quality sequences were recovered for downstream analysis, with an average of 3,226 reads, ranging from 1,703 to 11,411 reads per sample.

Genera Abundance and Diversity of Oral Microecology Were Depleted During OSCC Development

FASTQ files generated in 16S rRNA gene were further analyzed using MicrobiomeAnalyst, a web-based online tool for comprehensive statistical analysis of microbiome data.

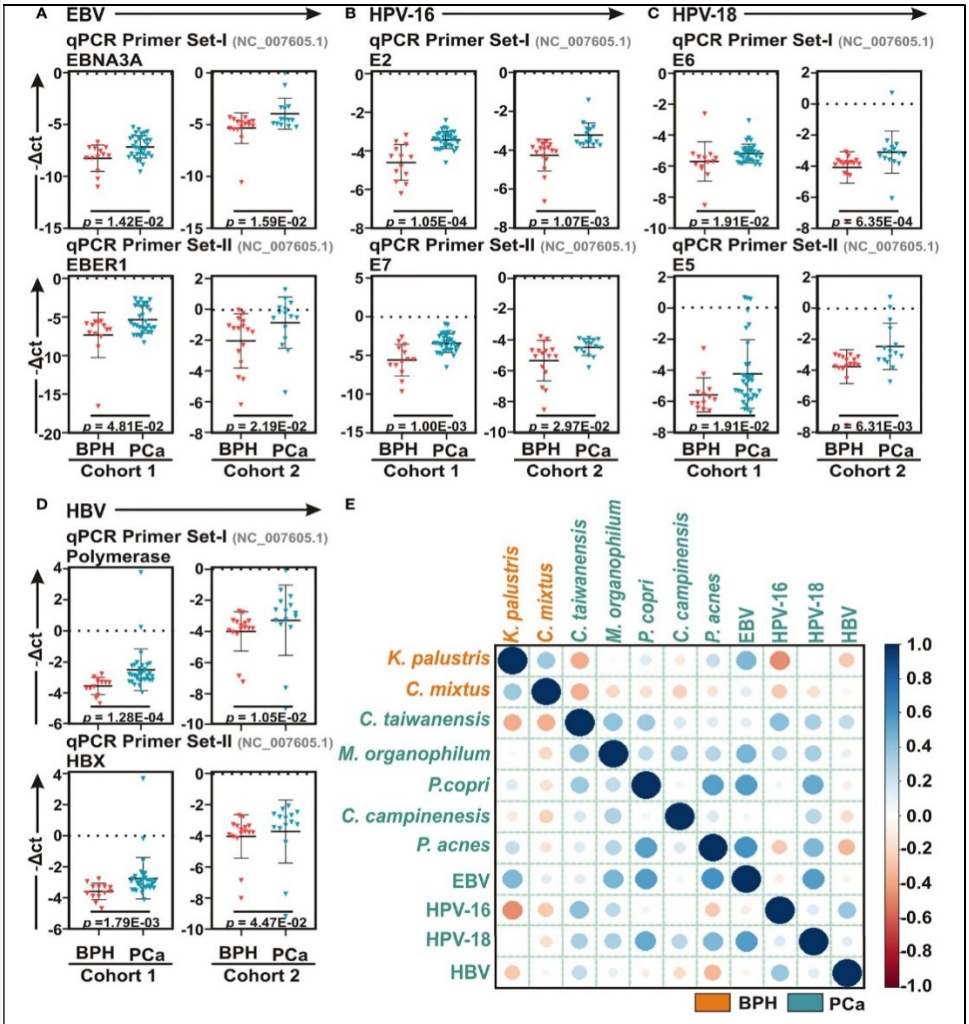

FIGURE 5: (A–D) Comparative qPCR data of (A) EBV, (B) HPV-16, (C) HPV-18 and (D) HBV among BPH and PCa samples. (E) Co-occurrence and co-exclusion analyses between most abundant bacteria and human oncogenic viruses in BPH and PCa lesions.

The average data of rarefaction plot demonstrated elevated genus richness in anatomically matched controls (samples 1–50) as compared to the paired OSCC lesions (samples 51–100), signifying a potential loss of certain genera during OSCC progression (Figure 6A). The OTU richness was further analyzed by calculating alpha diversity of oral microbiota, indicating the differences and similarities of the identified genera between the two sample groups (Figures 6B–E). The Observed genus ($p = 0.000463$) and Chao1 index ($p = 0.00101$) indicated that the OTU richness was significantly depleted in the OSCC samples. The diversity estimator Shannon index ($p = 0.00849$) indicated that relative diversity of bacterial genera was significantly decreased in cancerous lesions (Figures 6B–D respectively). A similar trend of depletion in relative diversity of bacterial genera in OSCC samples was also observed by employing Simpson index, although the data was not statistically significant ($p = 0.070680$) (Figure 6E).

Anatomically Matched OSCC and Normal Samples Comprised of Distinct Microbiome Composition

To estimate β-diversity, weighted UniFrac distances as well as Bray-Curtis dissimilarity metric were calculated from the OTU abundance and utilized in Principal Component Analysis (PCoA)(Figures 7A, B, respectively). The bacterial communities in the cancerous lesions and the anatomically matched controls clustered discretely, suggesting the overall structures of the bacterial communities in the groups were significantly different ($p < 0.002$ and $p < 0.009$, respectively in two analyses) (Figures 7A, B).

A 'Random Forest' algorithm was applied to further confirm the difference in bacterial community among the OSCC samples and anatomically matched healthy controls (Figure 7C). In the total of 50 OSCC samples, 32 samples revealed unique and 18 samples demonstrated overlapping genera; whereas in case of paired 50 normal tissue samples, 38 samples exhibited unique and 12 samples showed overlapping genera (Figure 7C).

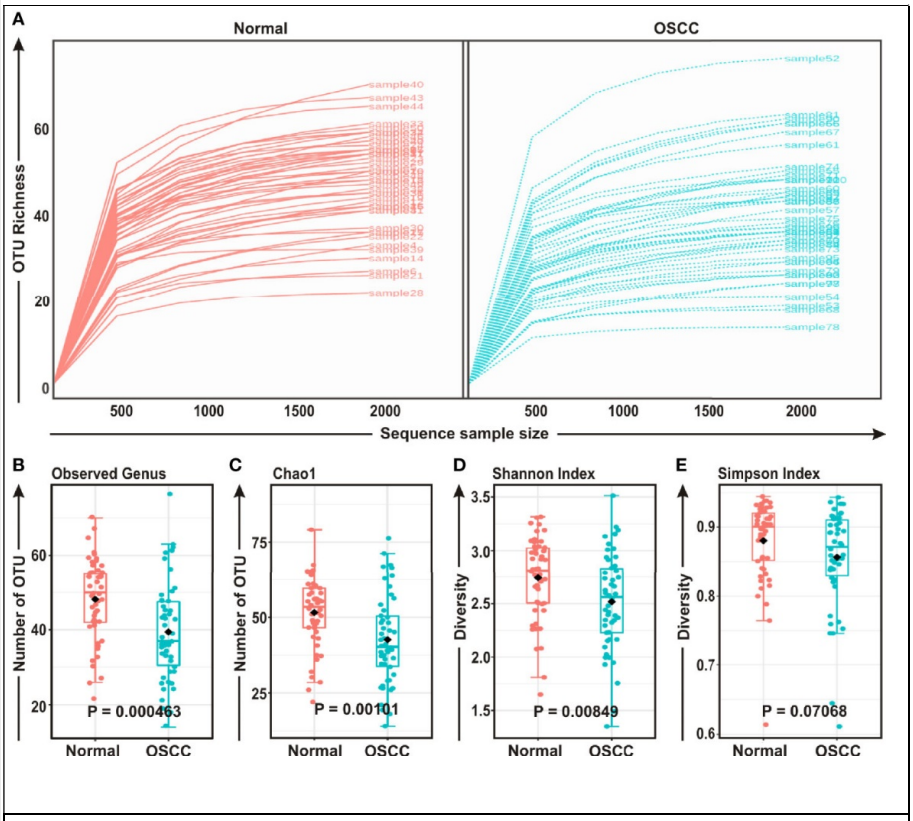

FIGURE 6: Comparison of oral microbiome compositions (A) Rarefaction analysis of bacterial 16S rRNA gene. (B–E) Box-Whisker plots of OUT abundance.

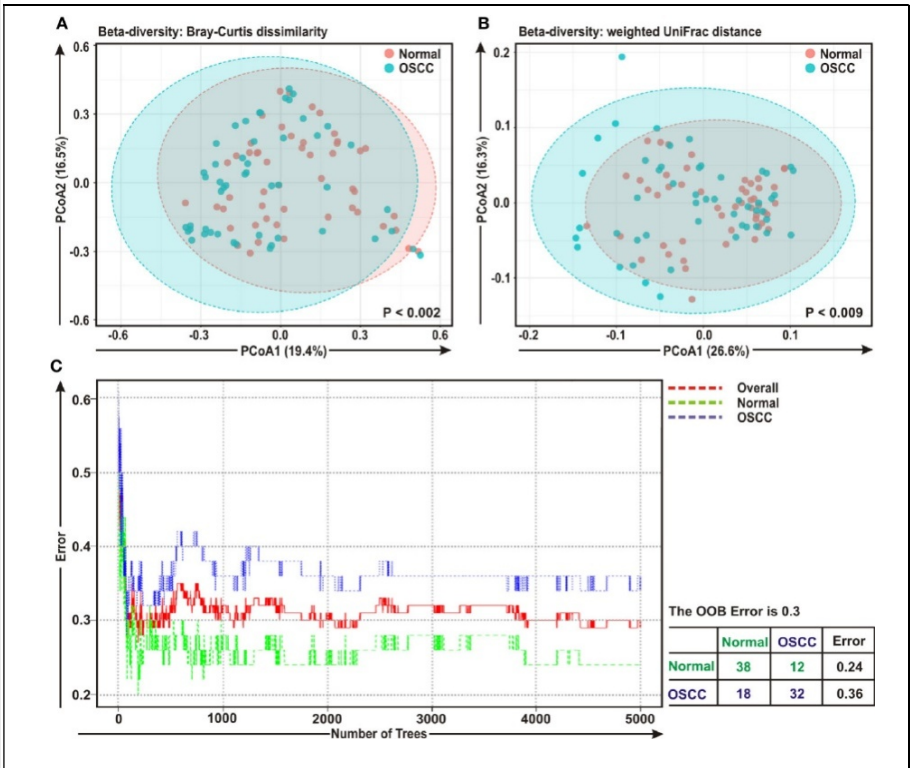

FIGURE 7: Beta diversity analyses among cancerous and normal samples. Principal Component Analysis (PCoA) plots based on (A) Bray-Curtis dissimilarity and (B) weighted UniFrac distance. (C) The error plots identified from random forest classification analyses.

Phylogenetic Analysis Revealed Variations Among Common and Distinct Taxa in OSCC Lesions and Anatomically Matched Healthy Controls:

The bacterial communities in the OSCC lesions and the anatomically matched healthy controls were first analyzed at phylum level (Figures 8A, B). The top five most abundant phyla comprised of 97.3% and 93% of all sequences in matched controls and OSCC lesions, respectively (Figure 8B). Firmicutes was the most abundant phylum in all samples, accounting for 36.1% of sequences in matched controls and 30.5% in OSCC lesions.

While abundance of phyla including Firmicutes, Proteobacteria and Actinobacteria were reduced, Fusobacteria, Bacteroidetes, Epsilonbacteraeota, and Spirochaetes were elevated in OSCC lesions compared to normal healthy controls (Figures 8A, B).

Of all genera detected, 137 taxa were found common in all samples, while 26 and 29 taxa were distinctly present OSCC lesions and anatomically matched controls, respectively. The shared genera among OSCC lesions and normal samples collectively represented more than 97.0% of all detected sequences in oral microbiota.

In order to further determine the differential presence and abundance of oral microbial community a phylogenetic tree was generated up to class level by MicrobiomeAnalyst (Figure

8C). The result demonstrated that Rudrobacteria class under phylum Actinobacteria was exclusively present in normal samples, whereas Deinococcus phylum was exclusively present in OSCC lesions (Figure 8C).

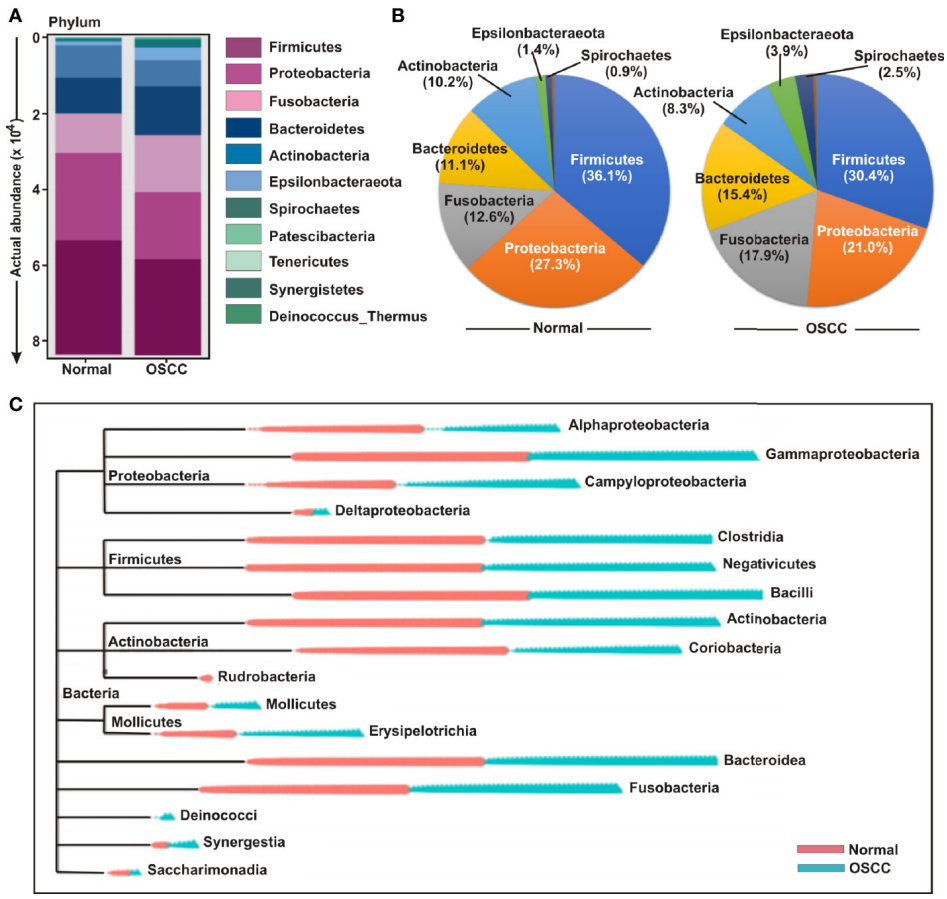

FIGURE 8: Composition of bacterial communities (A) Actual and (B) relative abundance of bacterial communities at phylum level. (C) Phylogenetic tree with operational taxonomic unit (OTU) abundances.

A number of bacterial genera under the classes of Gammaproteobacteria, Clostridia, Negativicutes, Bacilli, Actinobacteria, Bacteroidea and Fusobacteria demonstrated an elevated abundance in both groups (Figure 8C).

Enrichment Analysis Identified Unique Genera for OSCC and Adjacent Normal Tissue Samples:

A cladogram was generated using Galaxy, a web-based platform for bioinformatic analysis, to visualize significantly enriched bacteria taxa identified in OSCC lesions and anatomically matched adjacent control tissue samples (Figure 9A). Linear Discrimination Analysis (LDA) Effect Size (LEfSe) algorithm allows identifying high dimensional biomarkers among multiple study groups. To identify the distinguishing taxa within OSCC lesions and matched controls, we applied LEfSe method (Figures 9B, C). At the genus level, 22 taxa significantly declined in the OSCC lesions as compared to the healthy controls (Figure 9C). In contrast, only five taxa were significantly enriched in the OSCC lesions as compared to the anatomically matched control samples (Figure 9C).

Functional Prediction of Oral Microbiome Associated With OSCC Development:

To envisage oral microbial functions connected to the development of OSCC, we employed the PICRUSt and accordingly KEGG pathways were generated specific for OSCC lesions and anatomically matched healthy controls (Figure 10). The LEfSe outputs demonstrated that function related to nucleotide synthesis and maintaining the fundamental functions of a cell such as pyrimidine and purine metabolism, DNA repair and recombination proteins, DNA replication, transcription machinery, amino and nucleotide sugar metabolism, protein translation related function such as ribosome, amino acid related enzymes, aminoacyl tRNA biosynthesis, peptidases, as well as peptidoglycan biosynthesis were associated with the progression of OSCC (Figure 10A). In contrast, flagellar assembly, butanoate metabolism, secretion system, bacterial motility proteins, two-component system and ABC transporters, were inversely associated with OSCC

development (Figure 10A). PCoA analyses also demonstrated that the predicted functions of bacterial compositions in two groups – OSCC lesions and the anatomically matched controls were significantly clustered ($p < 0.05$) (Figure 10B).

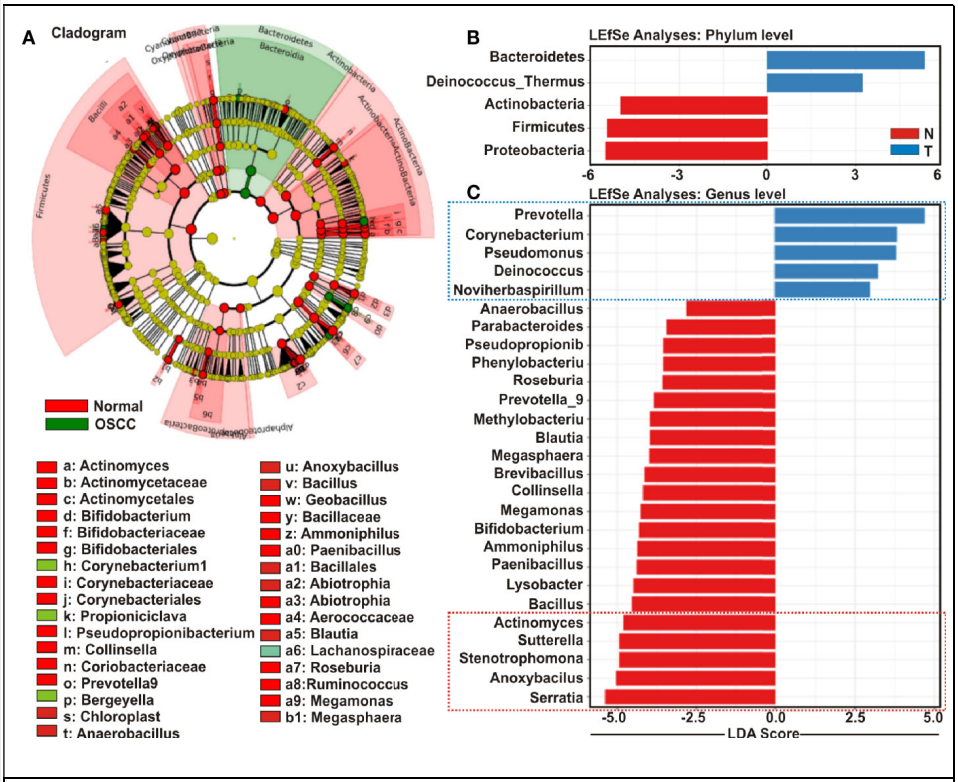

FIGURE 9: (A) Cladogram for phylogenetic relation of Normal and OSCC genus. LDA scores showed significant bacterial differences within groups OSCC and clinically normal counterparts at the (B) phylum level and (C) genus level.

Quantitative PCR Profile of Oncogenic Viruses Revealed Significant Association of HPV16 With OSCC Development:

To assess the potential involvement of viral etiology in our samples we designed real time PCR primers against EBV encoded EBNA3A oncogene along with two high risk HPV isotypes HPV-16 encoded E2 oncogene and HPV-18 encoded E6 oncogene and subsequently employed in quantitative PCR (qPCR) analyses (Figures 11A–C). The housekeeping gene human GAPDH gene was utilized as control. A higher negative -ΔCt indicated

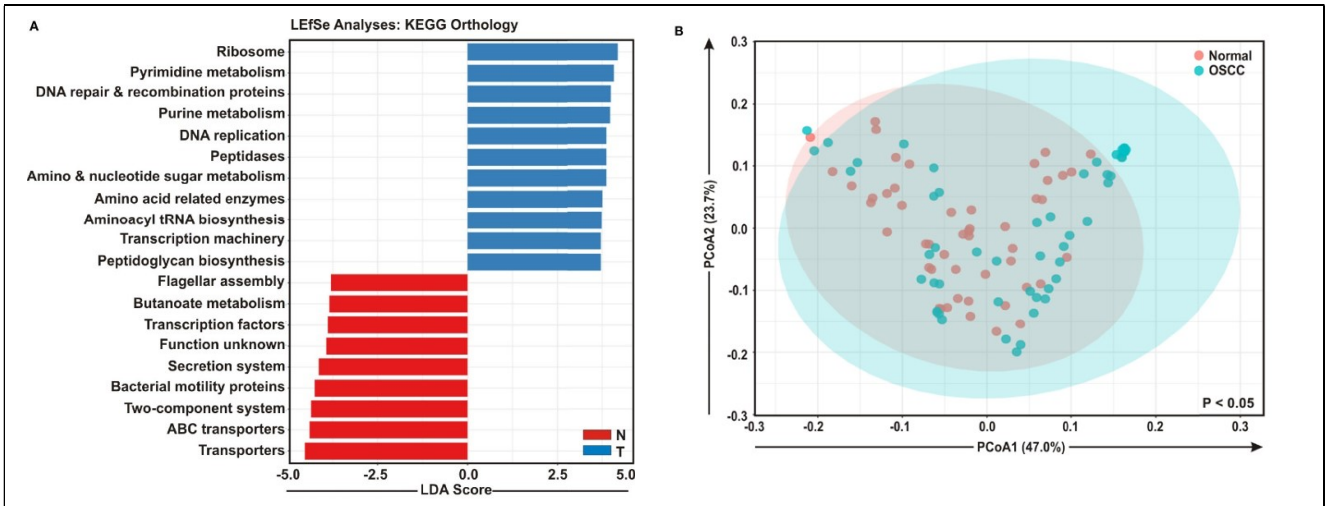

FIGURE 10: (A) Linear Discrimination Analysis (LDA) scores predicting gene function enriched among two different groups. (B) Principal Component Analysis (PCoA) of bacterial functions associated OSCC lesions and contralateral matched controls.

elevated presence of the virus in the sample as detected by specific primer set targeting specific viral gene. Our results clearly demonstrated that only HPV-16 ($p = 0.004$) was significantly associated with OSCC lesions as compared to the control tissue sections (Figure 11A, 11B, 11C). Altogether these oncogenic viruses might regulate the onset as well as progression of oral cavity oncogenesis and thereby demands their detection along with bacterial dysbiosis.

Next, the co-occurrence and co-exclusion patterns of these oncogenic viruses with the 27 most abundant bacterial genera identified in LEfSe analyses in each group of OSCC lesions and contralateral matched controls were further investigated (Figure 11D). Overall, there was no negative correlation found in our analyses. In matched normal controls, Prevotella_9 was found to be positively correlated with a number of bacterial genera.

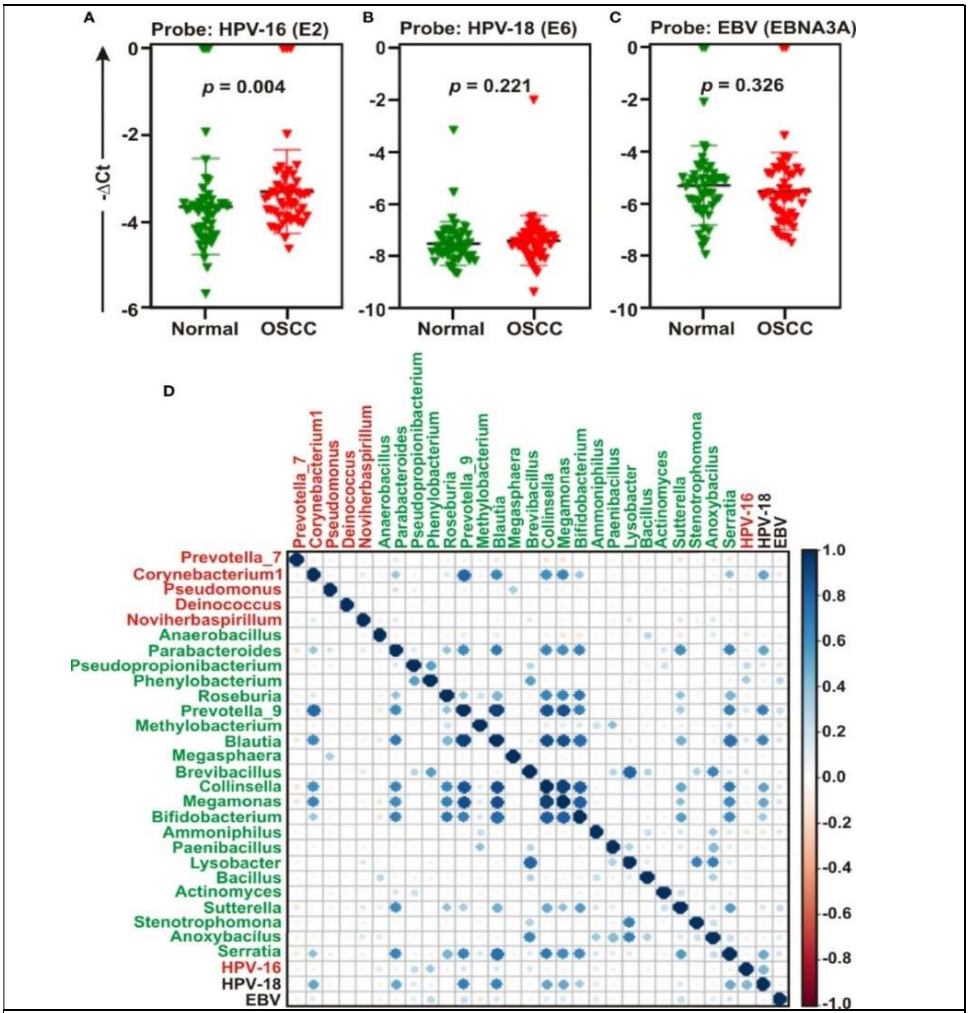

FIGURE 11: Comparative qPCR data of (A, B) high risk human papilloma viruses (HPVs) - (A) HPV-16 (B) HPV-18 and (C) Epstein-Barr virus (EBV). (D) Pearson correlations among human oncogenic viruses and the top 27 most abundant bacterial genera.

Discussion

A growing body of evidence indicates microbial infection as one of the predominant risk factors for both PCa and OSCC development[23,24,25,26]. In addition, several studies have evidently documented microbiome dysbiosis associated with these disease progressions[27,28,29,30]. Despite the increasing incidence of PCa and OSCC[31,32,33], to date, there are hardly any reports describing the microbial dysbiosis associated with these two cancers development among Indian patients. Herein, we identified specific microbial signatures, in both of these cancers from patients in the Eastern region of India.

Conclusion regarding Prostate cancer

Our study demonstrated significant enrichment of the steroid hormone biosynthesis pathway in PCa as compared to BPH samples. *Prevotella copri* is a frequent resident of the GI microbiome, and its higher prevalence has been consistently reported in non-westernised populations as well as patients with

systemic autoimmune disease[34,35]. Interestingly, our 16S rRNA sequencing results also demonstrated that *P. copri* was the most abundant species in both BPH and PCa lesions indicating potential involvement of the GI microbiota in PCa development, possibly through modulating the inflammatory response. Intriguing evidence is emerging that indicates a potential association of several human oncogenic viruses with PCa[36,37,38,39,40,41]. Moreover, the frequency of co-infection of EBV and high-risk HPV-18 has been shown to be significantly higher in PCa. In agreement to this, our results also demonstrated that EBV was strongly correlated with HPV-18 but not with HPV-16 in PCa samples. Our study for the first time demonstrated that HBV infection was significantly associated with PCa development. Importantly, a number of studies suggested a potential link between AR-signalling and HBV infection. It has been demonstrated that HBV oncoprotein HBx induces AR-responsive gene expression in an androgen concentration dependent manner[42,43]. It would be fascinating to investigate the role of HBV infection in PCa progression specially on castration-resistant PCa (CRPC) where AR signaling axis target is not sufficient for PCa therapy. Our study also clearly revealed that *C. taiwanensis*, *M. organophylum* followed by *P. acnes* and *P. copri* were the most correlated bacterial species with the infection status of EBV and HPV-18 and, to lesser extents, HPV-16 and HBV in PCa tissue specimens.

In sum, our study demonstrated that dysbiosis of tissue specific microbiome is directly linked to prostate health and disease. Future in-depth investigations with larger sample cohorts of different socioeconomic and ethnic backgrounds are required to discern whether the microbiota and/or their metabolites can be considered as novel biomarkers and therapeutic targets for PCa in Indian scenario.

Conclusion regarding Oral cancer

In addition, in line with the previous studies[44,45,46] *Firmicutes* was also found as the most abundant phylum in overall oral microbiome in our study. Of the significantly elevated genera in cancerous lesions, *Prevotella* and *Pseudomonus* were previously shown to be highly abundant in OSCC samples. Our results indicated presence of a unique genus *Deinococcus,* although relatively less abundant, only in cancerous lesions. Species of the *Deinococcus* genus are recognised for their extreme resistance to ionising radiation, oxidative stress, and other damaging conditions[47]. Although a number of earlier studies indicated the presence of the *Deinococcus* genus, the precise role of the members of this genus in OSCC is yet to be defined.

Nucleotide metabolism is an important pathway that provides purine and pyrimidine molecules for DNA replication, RNA biogenesis, as well as cell bioenergetics. Increased nucleotide metabolism represents a hallmark of cancer[48]. Apart from nucleotide metabolisms, several critical pathways like DNA repair, recombination, protein synthesis and transcription machineries are frequently altered in tumor cells[49,50,51]. In agreement to this, our PICRUSt analyses showed that functions related to nucleotide metabolisms as well as basic cell functions like DNA repair and replication and functions related to mRNA translation, including the ribosome, amino acid related enzymes, aminoacyl tRNA biosynthesis, and peptidases, were significantly linked to OSCC development. Herein we identified specific microbial signature associated with OSCC development. However, the current study has several limitations. Our study showed that the sequencing depth was not adequate to precisely classify the oral microbiota at the species level. Another limitation of this study was relatively smaller sample size. In future, larger sample size with distinct cancer stages among population in different regions and different socioeconomic background would be highly preferable. Altogether, longitudinal research activities are greatly demanded to explore the functional implications of the oral microbiota in terms of diagnosis and risk assessment of OSCC development, as well as potential expansion of current therapeutic strategies to restore the health of the oral ecosystem.

Table 1: Clinical characteristics of discovery and validation cohorts of Prostate samples.

	Sample ID	Age (yrs.)	Sample types	PSA	Gleason score	Family history	Smoking Habit
	Discovery Cohort (Cohort - 1)						
BPH* (N =13)	75	61	Biopsy	2.92	NA	No	Yes
	82	68	Biopsy	6	NA	No	Yes
	87	70	Biopsy	8	NA	No	Yes
	89	71	Biopsy	ND	NA	No	Yes
	96	56	Biopsy	4.27	NA	No	Yes
	108	60	Biopsy	12.9	NA	No	Yes
	122	60	Biopsy	0.19	NA	No	No
	128	46	Biopsy	26.69	NA	No	No
	117	60	Biopsy	0.553	NA	No	No
	119	60	Biopsy	16.97	NA	No	No
	102	77	Biopsy	8	NA	No	No
	103	80	Biopsy	51.81	NA	No	No
	113	78	Biopsy	25.34	NA	No	No
PCa# (N = 33)	38	63	Biopsy	19.2	4+3=7/10	No	Yes
	183	62	Biopsy	>1000	4+3=7/10	No	No
	143	50	Biopsy	162	3+4=7/10	No	Yes
	139	78	Biopsy	982	4+4=8/10	No	Yes
	58	58	Biopsy	145	3+4=7/10	No	Yes
	154	72	Biopsy	12.5	4+5=9/10	No	No
	92	52	Biopsy	157.4	3+4=7/10	No	Yes
	16	65	Biopsy	498	4+5=9/10	No	Yes
	135	63	Biopsy	3.9	5+4=9/10	No	No
	198	62	Biopsy	NA	3+4=7/10	No	No
	150	56	Biopsy	90	3+4=7/10	No	No
	100	40	Biopsy	980	5+5=10/10	No	No
	180	78	Biopsy	1777	3+4=7/10	No	No
	224	64	Biopsy	19.8	3+3=6/10	No	No
	93	62	Biopsy	57.08	3+3=6/10	No	Yes
	333	80	Biopsy	18.6	3+3 =6/10	No	No
	206	60	Biopsy	47.5	5+3=8/10	No	No
	131	71	Biopsy	68.4	4+3=7/10	Father	Yes
	202	70	Biopsy	0.8	4+4=8/10	No	No
	70	73	Biopsy	78.5	3+3=6/10	No	Yes
	134	74	Biopsy	85.75	3+3=6/10	No	Yes
	148	55	Biopsy	30.4	4+5=9/10	No	No
	327	67	Biopsy	>100	4+3 =7/10	No	No
	141	71	Biopsy	29.83	3+4=7/10	No	No
	187	65	Biopsy	40.2	3+4=7/10	No	No
	189	75	Biopsy	42	4+3=7/10	No	No
	31	67	Biopsy	64	4+4=8/10	No	No
	73	72	Biopsy	78.5	4+3=7/10	No	No
	324	63	Biopsy	4.46	3+4=7/10	No	No
	176	90	Biopsy	35.5	4+4=8/10	No	No
	151	60	Biopsy	52	4+3=7/10	No	No
	24	45	Biopsy	149	4+4=8/10	No	Yes
	157	56	Biopsy	28.48	3+4=7/10	No	No
Validation Cohort (Cohort - 2)							
BPH*	8	85	Biopsy	10	NA	No	No

(N =16)	15	55	Biopsy	14.29	NA	No	Yes
	20	75	Biopsy	8.03	NA	No	No
	21	72	Biopsy	ND	NA	No	No
	22	72	Biopsy	27.32	NA	No	No
	59	52	Biopsy	12	NA	No	Yes
	62	75	Biopsy	584	NA	No	Yes
	74	56	Biopsy	12.3	NA	No	No
	77	84	Biopsy	13.6	NA	No	No
	80	65	Biopsy	9.94	NA	No	Yes
	81	55	Biopsy	4.9	NA	No	Yes
	99	59	Biopsy	12.9	NA	No	Yes
	106	65	Biopsy	7.84	NA	No	Yes
	108	60	Biopsy	12.9	NA	No	Yes
	124	65	Biopsy	46.59	NA	No	No
	138	75	Biopsy	10.4	NA	No	Yes
PCa# (N =13)	33	70	Biopsy	145	4+3=7/10	No	Yes
	35	83	Biopsy	ND	4+4=8/10	No	Yes
	55	65	Biopsy	36.2	4+3=7/10	No	Yes
	319	78	Biopsy	5.1	4+4=8/10	No	-
	321	61	Biopsy	44.5	3+4=7/10	No	-
	339	49	Biopsy	45.93	4+4=8/10	No	-
	283	85	Biopsy	9.069	4+3=7/10	No	-
	305	64	Biopsy	36	3+3=6/10	No	-
	306	54	Biopsy	563	4+3=7/10	No	-
	268	67	Biopsy	56.7	4+4=8/10	No	-
	265	65	Biopsy	ND	5+5=10/10	No	-
	263	71	Biopsy	ND	ND	No	-
	262	56	Biopsy	16.3	3+4=8/10	No	-
	257	74	Biopsy	200	4+4=8/10	No	-
	207	72	Biopsy	88.7	3+3=6/10	No	-

Table 2: Clinical characteristics of OSCC patients with match control sample

Sl No.	Sample ID	Age	Sex	Clinical Diagnosis	Histopathological Diagnosis	Oral Habits
1	S1	37	M	Oral Carcinoma	Well differentiated Squamous Cell Carcinoma	Pan masala (20-30/day) for 7-8 years
2	S2	60	F	Malignancy with Verrucous Carcinoma	Well differentiated Squamous Cell Carcinoma	Pan masala chewing 4-5 times/day , areca nut chewing daily 4-5 packets for last 35 years.
3	S3	54	F	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
4	S4	36	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Gutkha (10-15 sachet/day for 2 years)
5	S5	62	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan masala chewing (10-12packets/day)since last 30 years
6	S6	70	F	Verrucous Carcinoma	Well differentiated Squamous Cell Carcinoma	Pan, khaini, tobbaco chewing for more than 20 years
7	S7	48	M	Tongue Carcinoma	Well differentiated Squamous Cell Carcinoma	No Habits
8	S8	44	M	Malignancy(recurrence case of maligancy)	Well differentiated Squamous Cell Carcinoma	Pan masala and jarda, but patient had left habit 2 years back when he was diagnosed with malignancy
9	S9	80	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Snuff rubbing daily for than 25 years
10	S10	42	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Gurakhu twice a day For more than 20 years
11	S11	63	M	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
12	S12	45	M	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
13	S13	35	M	OSMF with Malignancy	Well differentiated Squamous Cell Carcinoma	Khaini 8-10 times a day for more than 20 years, areca nut chewing and cigarette smoking occasionally
14	S14	50	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan chewing
15	S15	60	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan chewing 15 times a day
16	S16	62	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Bidi smoking 1 packet/day for more than 35 days
17	S17	55	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Khaini chewing for more than 30 years
18	S18	50	F	Maligncy	Well differentiated Squamous Cell Carcinoma	Placing 'Gurakhu' at the site of the lsion 3-4 times / day for 15 mins / placement
19	S19	45	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan chewing (15/day)
20	S20	55	F	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
21	S21	40	M	Tongue Carcinoma	Well differentiated Squamous Cell Carcinoma	Pan masala & jarda Supari chewing
22	S22	65	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Tobacco chewing for 30 years, quitted last 10 years
23	S23	73	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Gudaku brushing
24	S24	60	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Bidi, Quitted last 5-6 years
25	S25	60	M	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits

26	S26	60	M	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
27	S27	32	M	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
28	S28	45	M	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
29	S29	60	M	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
30	S30	40	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan masala chewing with supari occassionally since 20 years
31	S31	61	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan and Khaini
32	S32	45	F	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
33	S33	47	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Gutka (10-15 packets/day since 15 years) and alcohol 120 ml / day since 6-7 years
34	S34	30	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan,tobacco(6/day)
35	S35	59	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Tobacco chewing since 10years
36	S36	44	M	Erythroplakia	Well differentiated Squamous Cell Carcinoma	Pan (4-5/day) and bidi (4/day) for last 20 years
37	S37	38	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Khaini (20 times/day) and pan with jarda (6-8/day) for ~15 years
38	S38	60	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan masala 8-10 packets/day &bidi 1packet/day smoking since more than 25 years (left habitsince 2 months)
39	S39	62	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan chewing 2-3/day
40	S40	65	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan chewing 8-9/day for more than 25 years
41	S41	65	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Nassi intake
42	S42	41	M	Verrucous Carcinoma	Well differentiated Squamous Cell Carcinoma	No Habits
43	S43	52	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Gurakhu rubbing regularly for more than 15 years
44	S44	56	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan masala chewing 7-8 packets daily
45	S45	42	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Bidi smoking (7-8/day), Khaini (3-4packets /day), both for more 20 years
46	S46	40	M	Verrucous Hyperplasia	Well differentiated Squamous Cell Carcinoma	Khaini for more than 15 years
47	S47	55	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Bidi smoking 1 packet (25 pieces) per day for more than 20 years
48	S48	57	M	Malignancy	Well differentiated Squamous Cell Carcinoma	Khaini And pan masala 4-6/day for more than 25 years
49	S49	65	M	Malignancy	Well differentiated Squamous Cell Carcinoma	No Habits
50	S50	62	F	Malignancy	Well differentiated Squamous Cell Carcinoma	Pan chewing along with tobacco

Table3. Primers for real time quantitative PCR analyses.

Name of the species	Gene Segment	Accession ID	Primers
Human papillomavirus 16 (HPV-16)	E2	NC_001526.4	5'-CAGACGACTATCCAGCGACC-3' 5'-GCAGTGAGGATTGGAGCACT-3'
	E7		5'-GACAAGCAGAACCGGACAGA-3' 5'-GAGAACAGATGGGGCACACA-3'
Human papillomavirus 18 (HPV-18)	E6	NC_001357.1	5'-AAGCTCAGCAGACGACCTTC-3' 5'-CCCTCCCCGTCTGTACCTTA-3'
	E7		5'-CCGAACCACAACGTCACACA-3' 5'-GGAATGCTCGAAGGTCGTCT-3'
Epstein-Barr virus (EBV)	EBNA3A	NC_007605.1	5'-GCCCTGGATGACAACATGGA-3' 5'-CAGGTGGGCATCTTCTGCTT-3'
	EBER1		5'-GGAGACGTGTGTGGCTGTAG-3' 5'-TTGACCGAAGACGGCAGAAA-3'
Hepatitis B virus (HBV)	Polymerase	NC_003977.2	5'-GATGTGTCTGCGGCGTTTTA-3' 5'-GCAACATACCTTGATAGTCCAGAAGAA-3'
	X protein		5'-CCCGTCTGTGCCTTCTCATC-3' 5'-GTCGGTCGTTGACATTGCTG-3'
JC Polyomavirus (JCV)	Jvgp4	NC_001699.1	5'-GGAGACGTGTGTGGCTGTAG-3' 5'-TGTCATGAGTTGCTTGCCCA-3'
BK Polyomavirus (BKV)	VP1	NC_001538.1	5'-CCCTTGGAAATGCAGGGAGT-3' 5'-CATTACCTGGGACTGGGCTG-3'
Human herpesvirus 8 (KSHV)	ORF27	NC_009333.1	5'-TGGCTCGAAAGCTTCATCGT-3 5'-GCACAGAAACAGGGCTAGGT-3'
Merkel cell polyomavirus (MCPyV)	VP1	NC_010277.2	5'-GTGCCATCCGTTCTGGAAGA-3' 5'-CGAACACCATGAGGAACCCA-3'
Human T-lymphotropic virus 1 (HTLV-1)	Env	NC_001436.1	5'-ACTGGACCCACTGCTTTGAC-3' 5'-CTAGCGTGGGAACAGGTGAC-3'
Hepatitis C virus (HCV)	polyprotein;protein F	NC_038882.1	5'-GAGCGGTCTACGCCTTCTAC-3' 5'-ATATGGCGACAGAGTCAGCG-3'

Cellvibrio mixtus	endo-beta-1,4-xylanase	Z48926.1	5'-CTCATCAGCCCAAAGCTCCA-3' 5'-CACGGCTTGTTCCGAGGTAT-3'
	MFS transporter		5'-GTACATCCTGCGGGTCAACA-3' 5'-TGACGCAACTGTCTCTGCAT-3'
Cupriavidus taiwanensis	Homospermidine synthase	NC_010528.1	5'-CCAGGTGATGCCAACGTCTA-3' 5'-TCGTAAAGTTCGGCCCAGTC-3'
	FAD-dependent oxidoreductase		5'-CACCTGCTTGTGGTTTTCCG-3' 5'-CCATCTCGACATCACTCCCG-3'
Prevotella copri	YgiQ family SAM protein	NZ_GG703852.1	5'-CTCCAGCGCCAAGACTCAGA-3' 5'-ACCCTTGAAACTGCCACCCA-3'
	Pyruvate:ferredoxin oxidoreductase		5'-AGAAGGGTGAGAAGGCGTTG-3' 5'-CCAGAGCAAGCACCAGAGAA-3'
Propionibacterium acnes	RuvX	NC_006085.1	5'-GGAAAAGCACGGATTGGCGT-3' 5'-ATAGCCTCGGTCAACTGGCG-3'
	TIGR03767 family metallophosphoesterase		5'-CAGCAGTCCCCAGTCAGAAG-3' 5'-GGTTTTCCCCAGGTAGACGG-3'
Cupriavidus campinensis	Acetyl-CoA C-acyltransferase	FGG12_RS01630	5'-CGAGATGTTGTAGCGCTTGG -3' 5'-GGAGATGAACCGCCACATGA -3'
	Isochorismatase family protein		5'-GGCAAAGCCGTTCACGAAAT -3' 5'-TGCATGAGGAAGTCCAGACC -3'
Kocuria palustris	ABC-F family ATP-binding protein	NZ_CP012507.1	5'-GGCCTGAAGCTCGGCTACTT-3' 5'-GGCCAGTCGGGTCTTCTCAC-3'
	Glutamate synthase large subunit		5'-GTCATGCCGTACTTCGTCCA-3' 5'-AACGACGAGCAGTAGATGCC-3'
Methylobacterium organophilum	Methanol dehydrogenase alpha subunit-like gene	EF562471.1	5'-TTTGCGGCACCCAACTG-3' 5'-CCGAAGATCGTCATCGTCCA-3'
	Formyltransferase/hydrolase complex subunit D		5'-CCTCGACTAACGACGCCTAC-3' 5'-GTAGTTGCCGGCCGTGAC-3'
Human	Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)	NG_007073.2	5'-AGCTGGCCCGATTTCTCCTC-3' 5'-ATGACTCAGCTTCTCCCGGC-3'

References

[1]Sfanos, K. S., Yegnasubramanian, S., Nelson, W. G., De Marzo, A. M. (2017). The Inflammatory Microenvironment and Microbiome in Prostate Cancer Development. *Nat. Rev. Urol* 15, 11–24. doi: 10.1038/nrurol.2017.167

[2] Nickel, J. C., Xiang, J. (2008). Clinical Significance of Nontraditional Bacterial Uropathogens in the Management of Chronic Prostatitis. *J. Urol* 179, 1391–1395. doi: 10.1016/j.juro.2007.11.081

[3] Chen, Y., Wei, J. (2015). Identification of Pathogen Signatures in Prostate Cancer Using RNA-Seq. *PLoS One* 10, e0128955. doi: 10.1371/journal.pone.0128955

[4] Cavarretta, I., Ferrarese, R., Cazzaniga, W., Saita, D., Luciano, R., Ceresola, E. R., et al. (2017). The Microbiome of the Prostate Tumor Microenvironment. *Eur. Urol* 72, 625–631. doi: 10.1016/j.eururo.2017.03.029

[5] Yow, M. A., Tabrizi, S. N., Severi, G., Bolton, D. M., Pedersen, J., Giles, G. G., et al. (2017). Characterisation of Microbial Communities Within Aggressive Prostate Cancer Tissues. *Infect. Agent Cancer* 12, 4. doi: 10.1186/s13027-016-0112-7

[6]Jain, S., Saxena, S., Kumar, A. (2014). Epidemiology of Prostate Cancer in India. *Meta Gene* 2, 596–605. doi: 10.1016/j.mgene.2014.07.007

[7]Deo PN, Deshmukh R. Oral microbiome: Unveiling the fundamentals. *J Oral MaxillofacPathol* (2019) 23(1):122–8. doi: 10.4103/jomfp.JOMFP_304_18JOMFP-23-122

[8] Curtis MA, Zenobia C, Darveau RP. The relationship of the oral microbiotia to periodontal health and disease. *Cell Host Microbe* (2011) 10(4):302–6. doi: 10.1016/j.chom.2011.09.008

[9]Preza D, Olsen I, Aas JA, Willumsen T, Grinde B, Paster BJ. Bacterial profiles of root caries in elderly patients. *J ClinMicrobiol* (2008) 46(6):2015–21. doi: 10.1128/JCM.02411-07

[10] Chhour KL, Nadkarni MA, Byun R, Martin FE, Jacques NA, Hunter N. Molecular analysis of microbial diversity in advanced caries. *J ClinMicrobiol* (2005) 43(2):843–9. doi: 10.1128/JCM.43.2.843-849.2005

[11]Tezal M, Sullivan MA, Hyland A, Marshall JR, Stoler D, Reid ME, et al. Chronic periodontitis and the incidence of head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* (2009) 18(9):2406–12. doi: 10.1158/1055-9965.EPI-09-0334

[12]Karpinski TM. Role of Oral Microbiota in Cancer Development. *Microorganisms* (2019) 7(1):20. doi: 10.3390/microorganisms7010020

[13]Malla MA, Dubey A, Kumar A, Yadav S, Hashem A, Abd Allah EF. Exploring the Human Microbiome: The Potential Future Role of Next-Generation Sequencing in Disease Diagnosis and Treatment. *Front Immunol* (2019) 9:2868:2868. doi: 10.3389/fimmu.2018.02868

[14]Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* (2010) 7(5):335–6. doi: 10.1038/nmeth.f.303

[15]Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* (2007) 73(16):5261–7. doi: 10.1128/AEM.00062-07

[16]Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* (2014) 42(Database issue):D633–42. doi: 10.1093/nar/gkt1244

[17]Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* (2018) 46(W1):W537–W44. doi: 10.1093/nar/gky379

[18]Basu S, Kumbier K, Brown JB, Yu B. Iterative random forests to discover predictive and stable high-order interactions. *Proc Natl AcadSci U.S.A.* (2018) 115(8):1943–8. doi: 10.1073/pnas.1711236115

[19]Kelly BJ, Gross R, Bittinger K, Sherrill-Mix S, Lewis JD, Collman RG, et al. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* (2015) 31(15):2461–8. doi: 10.1093/bioinformatics/btv183

[20]Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* (2013) 31(9):814–21. doi: 10.1038/nbt.2676

[21]Kanehisa M, Goto S. KEGG: kyotoencyclopedia of genes and genomes. *Nucleic Acids Res* (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27

[22]Zhao H, Chu M, Huang Z, Yang X, Ran S, Hu B, et al. Variations in oral microbiota associated with oral cancer. *Sci Rep* (2017) 7(1):11773. doi: 10.1038/s41598-017-11779-910.1038/s41598-017-11779-9

[23]Su SC, Chang LC, Huang HD, Peng CY, Chuang CY, Chen YT, et al. Oral microbial dysbiosis and its performance in predicting oral cancer. *Carcinogenesis* (2020) 4:bgaa062. doi: 10.1093/carcin/bgaa062

[24]Sfanos, K. S., Yegnasubramanian, S., Nelson, W. G., De Marzo, A. M. (2017). The Inflammatory Microenvironment and Microbiome in Prostate Cancer Development. *Nat. Rev. Urol* 15, 11–24. doi: 10.1038/nrurol.2017.167

[25]Sutcliffe, S., Nevin, R. L., Pakpahan, R., Elliott, D. J., Cole, S. R., De Marzo, A. M., et al. (2011). Prostate Involvement During Sexually Transmitted Infections as Measured by Prostate-Specific Antigen Concentration. *Br. J. Cancer* 105, 602–605. doi: 10.1038/bjc.2011.271

[26]Whitaker, N. J., Glenn, W. K., Sahrudin, A., Orde, M. M., Delprado, W., Lawson, J. S. (2013). Human Papillomavirus and Epstein Barr Virus in Prostate Cancer: Koilocytes Indicate Potential Oncogenic Influences of Human Papillomavirus in Prostate Cancer. *Prostate* 73, 236–241. doi: 10.1002/pros.22562

[27]Cavarretta, I., Ferrarese, R., Cazzaniga, W., Saita, D., Luciano, R., Ceresola, E. R., et al. (2017). The Microbiome of the Prostate Tumor Microenvironment. *Eur. Urol* 72, 625–631. doi: 10.1016/j.eururo.2017.03.029

[28]Yow, M. A., Tabrizi, S. N., Severi, G., Bolton, D. M., Pedersen, J., Giles, G. G., et al. (2017). Characterisation of Microbial Communities Within Aggressive Prostate Cancer Tissues. *Infect. Agent Cancer* 12, 4. doi: 10.1186/s13027-016-0112-7

[29]Banerjee, S., Alwine, J. C., Wei, Z., Tian, T., Shih, N., Sperling, C., et al. (2019). Microbiome Signatures in Prostate Cancer. *Carcinogenesis* 40, 749–764. doi: 10.1093/carcin/bgz008

[30]Massari, F., Mollica, V., Di Nunno, V., Gatto, L., Santoni, M., Scarpelli, M., et al. (2019). The Human Microbiota and Prostate Cancer: Friend or Foe? *Cancers (Basel)* 11, 459. doi: 10.3390/cancers11040459

[31]Basu B, Chakraborty J, Chandra A, Katarkar A, Baldevbhai JRK, Dhar Chowdhury D, et al. Genome-wide DNA methylation profile identified a unique set of differentially methylated immune genes in oral squamous cell carcinoma patients in India. *Clin Epigenet* (2017) 9:13. doi: 10.1186/s13148-017-0314-x314

[32]Byakodi R, Byakodi S, Hiremath S, Byakodi J, Adaki S, Marathe K, et al. Oral cancer in India: an epidemiologic and clinical review. *J Community Health* (2012) 37(2):316–9. doi: 10.1007/s10900-011-9447-6

[33]Jain, S., Saxena, S., Kumar, A. (2014). Epidemiology of Prostate Cancer in India. *Meta Gene* 2, 596–605. doi: 10.1016/j.mgene.2014.07.007

[34]Scher, J. U., Sczesnak, A., Longman, R. S., Segata, N., Ubeda, C., Bielski, C., et al. (2013). Expansion of Intestinal Prevotella Copri Correlates With Enhanced Susceptibility to Arthritis. *Elife* 2, e01202. doi: 10.7554/eLife.01202.028

[35]Tett, A., Huang, K. D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., et al. (2019). The Prevotella Copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host Microbe* 26, 666–679.e667. doi: 10.1016/j.chom.2019.08.018

[36]Samanta, M., Harkins, L., Klemm, K., Britt, W. J., Cobbs, C. S. (2003). High Prevalence of Human Cytomegalovirus in Prostatic Intraepithelial Neoplasia and Prostatic Carcinoma. *J. Urol* 170, 998–1002. doi: 10.1097/01.ju.0000080263.46164.97

[37]Das, D., Wojno, K., Imperiale, M. J. (2008). BK Virus as a Cofactor in the Etiology of Prostate Cancer in its Early Stages. *J. Virol.* 82, 2705–2714. doi: 10.1128/JVI.02461-07

[38]Lin, Y., Mao, Q., Zheng, X., Yang, K., Chen, H., Zhou, C., et al. (2011). Human Papillomavirus 16 or 18 Infection and Prostate Cancer Risk: A Meta-Analysis. *Ir J. Med. Sci.* 180, 497–503. doi: 10.1007/s11845-011-0692-6

[39]Chen, Y., Wei, J. (2015). Identification of Pathogen Signatures in Prostate Cancer Using RNA-Seq. *PLoS One* 10, e0128955. doi: 10.1371/journal.pone.0128955

[40]Smelov, V., Bzhalava, D., Arroyo Muhr, L. S., Eklund, C., Komyakov, B., Gorelov, A., et al. (2016). Detection of DNA Viruses in Prostate Cancer. *Sci. Rep.* 6, 25235. doi: 10.1038/srep25235

[41] Banerjee, S., Alwine, J. C., Wei, Z., Tian, T., Shih, N., Sperling, C., et al. (2019). Microbiome Signatures in Prostate Cancer. *Carcinogenesis* 40, 749–764. doi: 10.1093/carcin/bgz008

[42] Chiu, C. M., Yeh, S. H., Chen, P. J., Kuo, T. J., Chang, C. J., Yang, W. J., et al. (2007). Hepatitis B Virus X Protein Enhances Androgen Receptor-Responsive Gene Expression Depending on Androgen Level. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2571–2578. doi: 10.1073/pnas.0609498104

[43] Yang, W. J., Chang, C. J., Yeh, S. H., Lin, W. H., Wang, S. H., Tsai, T. F., et al. (2009). Hepatitis B Virus X Protein Enhances the Transcriptional Activity of the Androgen Receptor Through C-Src and Glycogen Synthase Kinase-3beta Kinase Pathways. *Hepatology* 49, 1515–1524. doi: 10.1002/hep.22833

[44] Nagy KN, Sonkodi I, Szoke I, Nagy E, Newman HN. The microflora associated with human oral carcinomas. *Oral Oncol* (1998) 34(4):304–8. doi: 10.1016/S1368-8375(98)80012-2

[45] Bolz J, Dosa E, Schubert J, Eckert AW. Bacterial colonization of microbial biofilms in oral squamous cell carcinoma. *Clin Oral Investig* (2014) 18(2):409–14. doi: 10.1007/s00784-013-1007-2

[46] de Andrade KQ, Almeida-da-Silva CLC, Coutinho-Silva R. Immunological Pathways Triggered by Porphyromonas gingivalis and Fusobacterium nucleatum: Therapeutic Possibilities? *Mediators Inflammation* (2019) 2019:7241312. doi: 10.1155/2019/7241312

[47] Pavlopoulou A, Savva GD, Louka M, Bagos PG, Vorgias CE, Michalopoulos I, et al. Unraveling the mechanisms of extreme radioresistance in prokaryotes: Lessons from nature. *Mutat Res Rev Mutat Res* (2016) 767:92–107. doi: 10.1016/j.mrrev.2015.10.001

[48]Siddiqui A, Ceppi P. A non-proliferative role of pyrimidine metabolism in cancer. *Mol Metab* (2020) 35:100962. doi: 10.1016/j.molmet.2020.02.005

[49] Dietlein F, Thelen L, Reinhardt HC. Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. *Trends Genet* (2014) 30(8):326–39. doi: 10.1016/j.tig.2014.06.003

[50]Bradner JE, Hnisz D, Young RA. Transcriptional Addiction in Cancer. *Cell* (2017) 168(4):629–43. doi: 10.1016/j.cell.2016.12.013

[51] Lee LJ, Papadopoli D, Jewer M, Del Rincon S, Topisirovic I, Lawrence MG, et al. Cancer Plasticity: The Role of mRNA Translation. *Trends Cancer* (2020) 7(2):134–45. doi: 10.1016/j.trecan.2020.09.005